

# On the minimal accuracy required for simulating self-gravitating systems by means of direct N-body methods

Simon Portegies Zwart

Leiden Observatory, Leiden University, PO Box 9513, 2300 RA, Leiden, The Netherlands

and

Tjarda Boekholt

Leiden Observatory, Leiden University, PO Box 9513, 2300 RA, Leiden, The Netherlands

Received \_\_\_\_\_; accepted \_\_\_\_\_

## ABSTRACT

The conservation of energy, linear momentum and angular momentum are important drivers for our physical understanding of the evolution of the Universe. These quantities are also conserved in Newton’s laws of motion under gravity (Newton 1687). Numerical integration of the associated equations of motion is extremely challenging, in particular due to the steady growth of numerical errors (by round-off and discrete time-stepping, Babadzhanyants (1981); Goodman et al. (1993); Hernquist et al. (1993); Makino (1994)) and the exponential divergence (Miller 1964; Urmitsky & Heggie 2009) between two nearby solution. As a result, numerical solutions to the general N-body problem are intrinsically questionable (Heggie & Hut 2003; Lyapunov & Walker 1994). Using brute force integrations to arbitrary numerical precision we demonstrate empirically that ensembles of different realizations of resonant 3-body interactions produce statistically indistinguishable results. Although individual solutions using common integration methods are notoriously unreliable, we conjecture that an ensemble of approximate 3-body solutions accurately represents an ensemble of true solutions, so long as the energy during integration is conserved to better than 1/10. We therefore provide an independent confirmation that previous work on self-gravitating systems can actually be trusted, irrespective of the intrinsic chaotic nature of the N-body problem.

## 1. Introduction

Newton’s law of gravitation is one of the fundamental laws in the Universe that holds everything together. Although formulated in the 17th century, scientists today still study

the consequences, in particular those of many-body systems, like the solar system, star clusters and the Milky Way galaxy. General analytic solutions to the  $N$ -body problem only exist for configurations with one mass, commonly referred to as  $N = 1$  solutions, and for two masses (equivalently named  $N = 2$ , Kepler (1609); Newton (1687)). Problems for  $N \rightarrow \infty$  can be reduced via Liouville’s theorem for Hamiltonian systems to the collisionless Boltzmann equation (Boltzmann 1868; Vlasov 1968, but see also Maxwell (1867)), and therefore analytic solutions for the global distribution function exist.

Solutions for  $N$  in between these two limits are generally realized by computer simulations. These so-called  $N$ -body simulations have a major shortcoming in that the solution to any initial realization can only be approximated. The main limiting factors in numerically obtaining a true solution include errors due to round-off and approximations both in the integration and in the time-step strategy (Kandrup 1996). These generally small errors are magnified by the exponentially sensitive dependence on the  $6N$ -dimensional phase-space coordinates, position and velocity (Miller 1964). As a consequence, the solution for a numerically integrated self-gravitating system of  $N$  masses diverges from the true solution. This error can be controlled to some degree by selecting a phase-space volume preserving or a symplectic algorithm (Wisdom & Holman 1991) and by reducing the integration time step (Aarseth & Lecar 1975; Hayes 2008). The latter however, cannot be reduced indefinitely due to the accumulation of numerical round-off in the mantissa, which is generally limited to 53 bits (64 bits in total, but 11 bits are reserved for the exponent, resulting in only about 15 significant digits). The exponential divergence subsequently causes this small error to propagate to the entire system on a dynamical time scale (Goodman et al. 1993), which is the time scale for a particle to cross the system once. The result of these errors together with the exponential divergence, is the loss of predicting power for a numerical solution to a self gravitating system with  $N > 2$  after a dynamical time scale. One can subsequently question the predicting qualities of  $N$ -body simulations

for self-gravitating systems, and therewith their usefulness as a scientific instrument.

We address this question for  $N = 3$  by brute-force numerical integration to arbitrary precision. The choice of  $N = 3$  is motivated by the realization that this represents the first fundamental irregular configuration with the smallest possible number of objects that cannot be solved analytically and cannot be addressed with collisionless theory. In addition, 3-body encounters form a fundamental and frequently occurring topology in any large  $N$ -body simulation, and therefore also drive the global dynamics of these larger systems.

## 2. Validation of the unrestricted precision integration

The divergence between two different, approximate solutions to the  $N$ -body problem can be quantified by the phase-space distance in the positions  $\mathbf{r}$  and velocities  $\mathbf{v}$  of the  $N$  particles (in dimension-less  $N$ -body units):

$$\delta^2 = \frac{1}{6N} \sum_{i=1}^N [(\mathbf{r}_A - \mathbf{r}_B)^2 + (\mathbf{v}_A - \mathbf{v}_B)^2]. \quad (1)$$

Values of  $\delta$  are obtained by comparing the configurations from solution  $A$  and solution  $B$  at any moment in time. Each star has a position and velocity in solution  $A$  and (generally) a different position and velocity in solution  $B$ . For each star we calculate its phase-space distance between the two solutions. By dividing by  $6N$ ,  $\delta$  can be thought of as the average difference per coordinate. The two different runs can be performed either with the same code at a different precision, or with two different codes, all having exactly the same initial realization. A value of  $\delta \gtrsim 0.1$  indicates that the results of the two simulations have diverged beyond recognition. We consider a solution to be converged to  $p$  decimal places when, for any time  $t > 0$ ,  $\delta < 10^{-p}$ . (In stable hierarchical few-body systems the value of  $\delta$  can vary substantially across the orbital phase (Dejonghe & Hut 1986), and one has to assure that temporary large deviations can diminish again at a later instant.)

To investigate the build-up of numerical errors and the corresponding exponential divergence, we developed an  $N$ -body solver for self-gravitating systems which solves the  $N$ -body problem to arbitrary precision. This code, named **Brutus** (Boekholt & Portegies Zwart, in preparation), is composed of a Bulirsch-Stoer integrator (Bulirsch & Stoer 1964), that preserves energy to the level of the Bulirsch-Stoer tolerance. This tolerance is a parameter which can be interpreted as the discretization error per integration step. The round-off error is controlled by choosing the word-length with which all floating point numbers in the computer code are represented. By decreasing the Bulirsch-Stoer tolerance and increasing the word-length, we can obtain solutions to the  $N$ -body problem to arbitrary precision.

We tested **Brutus** by adopting a 3-body system of identical particles, which are located on the vertices of an equilateral triangle, with initial velocities such that the orbits are on a circle around the center of mass (de Lagrange 1772). Because this system is intrinsically unstable, small perturbations in the position and velocity vectors cause the triangular configuration to dissolve quickly. The time at which this happens depends on precision. Using **Brutus** we can reach arbitrary precision, but in this validation experiment we stopped reducing the time step and increasing the word length once the energy was conserved up-to 75 decimal places, which is sufficient to demonstrate our point. For any pre-determined time of stability there is a combination of word-length and Bulirsch-Stoer tolerance for which **Brutus** converges. We define a solution to be converged when the first  $p$  decimal places become independent of the size of the time step and the word length. This is equivalent to saying that  $\delta$  is always below  $10^{-p}$ ; for  $p = 3$  (at least the first 3 digits have converged), then  $\delta < 10^{-3}$  at all times.

### 3. Results

Having established the possibility of integrating a self-gravitating  $N$ -body system to arbitrary precision we can study the reliability of  $N$ -body simulations in general. We limit ourselves to the problem of 3 bodies, by generating a database of different 3-body problems and solve them until a converged solution is achieved. The positions of the particles are taken randomly from a Plummer (1911) distribution and are either cold (zero kinetic energy) or virialized. In the cold case we assured ourselves that the mutual distances between the particles are initially comparable (within an order of magnitude). We performed runs with identical masses and with the masses in a ratio of 1:2:4. For each of the four selected ensembles of initial conditions we generated  $10^4$  random realizations. The masses and coordinates for these systems are specified in standard double precision, to assure that the double-precision calculations use exactly the same initial realizations as the arbitrary-precision calculations. Every initial condition is integrated using the leapfrog-Verlet (Verlet (1967), we adopted the implementation available in <http://nbabel.org>) and the 4th-order Hermite predictor-corrector scheme in a code called `ph4` (McMillan et al. 2012). (Both codes, `Brutus` and `ph4` are assimilated in the public AMUSE framework which is available at <http://amusecode.org>, Portegies Zwart et al. (2013)). The integration continues until the system has been dissolved into a permanent binary and a single escaper (Heggie 1975; Hut & Bahcall 1983). Dissolution is declared upon the first integral dynamical time upon which one particle is unbound, outside a sphere of 2 initial virial radii around the barycenter, and receding from the center of mass (Hut & Bahcall 1983). A particle is considered unbound if its kinetic energy in the center of mass reference frame exceeds the absolute value of its potential energy, which is more strict than adopted in Hut & Bahcall (1983). For a fraction of the simulations (see Fig. 3), the dissolution time turns out to be very long as the evolution consists of a sequence of ejections where a particle almost escapes, but then still returns to once again enter a 3-body

resonance. We therefore put a constraint on the integration time and use the fraction of long-lived systems as a measurable statistic. We obtain ensembles of solutions using the Hermite and leapfrog integrators with a time-step parameter  $\eta = 2^{-1}, 2^{-2}, \dots 2^{-11}$ . Here we adopted the definition for  $\eta$  given by Aarseth & Lecar (1975).

We subsequently recalculate each of these initial realizations with **Brutus** using the same tolerance. In subsequent calculation we systematically reduce the time-step size and increase the word-length until we obtain a converged solution (as we discussed in §2 for  $p = 3$ ) for every realization of the initial conditions. This converged solution is then compared to the earlier simulations performed with the Hermite and leapfrog integrators.

We now have three solutions for each initial realization of the 3-body problem, one of which is the converged solution. We compare the three solutions for the time of dissolution, the semi-major axis (or equivalently the reciprocal of the orbital energy) of the surviving binary, its eccentricity (equivalent to the angular momentum) and the escaper’s velocity and direction.

In Fig. 1 we individually compare the time to dissolution for a certain initial realization as given by the Hermite integrator and the converged solution as given by **Brutus**. About half of the individual Hermite solutions lie along the diagonal representing the accurate solutions. The other half is scattered around the diagonal. These solutions have diverged away from the converged solution, producing a binary and an escaper with completely different properties. For dissolutions within  $\sim 10$  dynamical times, there is insufficient time for the solution to diverge and the results of the various numerical methods are consistent. But once the Hermite or leapfrog solutions have diverged away from the converged solution the entire parameter space of the numerical experiment is sampled. A similar statement holds when instead of comparing the dissolution time, we compare the properties of the binaries or the escapers.

In Fig. 2 we present the cumulative distribution function of the difference between the time to dissolution of the Hermite and **Brutus** calculations:  $dt_{\text{dissolve}} = t_{\text{Hermite}} - t_{\text{Brutus}}$  for three different values of  $\eta = 2^{-2}$ ,  $\eta = 2^{-3}$  and  $\eta = 2^{-9}$ . The differences for  $\eta \leq 2^{-3}$  are symmetric around the origin with a dispersion of  $\sim 70$   $N$ -body time units, but for  $\eta \geq 2^{-2}$  it is not symmetric. The distributions in the differences in semi-major axis, eccentricity and the direction of the escaper (polar and azimuthal angle with respect to the binary plane) at the time we stop the experiment for  $\eta = 2^{-3}$  down to  $\eta = 2^{-11}$  are symmetric with respect to the origin. The global distributions are statistically indistinguishable via a Kolmogorov-Smirnov test. We empirically determine that for a value of the time-step parameter  $\eta = 2^{-3}$  the majority of the ensemble conserves energy to better than 1/10.

In Fig. 3 we present the fraction of undissolved systems in time. The colored symbols give the converged solutions, whereas the curves give the results obtained using the Hermite integrator. The two solutions for each ensemble of initial realizations for  $\eta \leq 2^{-3}$  (as well as those obtained with the leapfrog integrator, not shown) are statistically indistinguishable after comparing  $10^4$  realizations of the initial conditions. The distributions obtained using  $\eta \geq 2^{-2}$  are not symmetric.

The duration of stability was studied as a function of accuracy by Urminsky (2010) using the Sitnikov problem (Moser 1973). They found that the remaining time for the system to stay bound depends on the integration accuracy. Our simulations did not reveal this effect, because we study systems that dissolve on a much shorter time-scale.

## 4. Conclusions

The properties of the binary and the escaper of a three-body system can be described in a statistical way. This is consistent with the findings in previous analytic (Monaghan 1976)



and numerical (Valtonen et al. 2004) studies. This behavior was named quasi-ergodicity by Monaghan (1976). We confirm that this behavior remains valid also for converged 3-body solutions.

Based on the symmetry of the distribution in dissolution times (see Fig. 2), the final parameters of the binary and escaper, as well as the consistency of the mean and median values of the inaccurate simulations when compared to the converged solution (see Fig. 2 and Fig. 3) we argue that global statistical distributions are preserved irrespective of the precision of the calculation as long as energy is preserved to better than 1/10th of the initial energy of the system. Although we have tested only three algorithms for solving the equations of motion we conjecture that the statistical consistency may be preserved also for some other direct  $N^2$  methods, and these may also require that energy and angular momentum are preserved to  $\leq 1/10$ th. If such direct  $N$ -body methods comply to the same statistical behavior for collision-less ( $N \gg 3$ ) systems, it will be interesting to investigate how also other –non- $N^2$ – algorithms, like the hierarchical tree-method (Barnes & Hut 1986) or particle-mesh methods (Hockney & Eastwood 1988) behave in this respect.

In studies of self-gravitating systems which adopt the 4th order Hermite integrator, energy and angular momentum are generally conserved up to  $\lesssim 10^{-6}$  per dynamical time. Only those simulations in which this requirement is met are often considered reliable and suitable for scientific interpretation. Proof for this seemingly conservative choice has never been provided, and it is unknown whether or not the numerical error and the exponential divergence are not preventing certain parts in parameter space to be accessed, or new physically inaccessible parts in parameters space being explored. We argued that for the resonant 3-body problem the error made during the integration of the equations of motion poses no problem for obtaining scientifically meaningful results so long as energy is conserved to better than about one-tenth of the initial total energy of the system. In that

case resonant 3-body interactions should be treated as an ensemble average, and individual results only contribute statistically.

By means of numerical integration until a converged solution is obtained we find that the statistical properties of the binary and the escaper resulting from a 3-body resonant encounter are deterministic. This behavior is not guaranteed to propagate to larger  $N$  (see also Quinlan & Tremaine (1992));  $N > 3$  requires independent testing, because these introduce more complex solutions in the form of, for example binary-binary outcomes and hierarchical triples. The more extended parameter space for increasing  $N$  from 3 to  $N = 4$  is quite dramatic, in particular for solving the system until a converged solution is reached.

**Acknowledgements** We would like to thank Douglas Heggie for many in-depth discussions, but also Alice Quillen, Piet Hut, Jun Makino, Steve McMillan, Vincent Icke and Inti Pelupessy for discussions and comments on the manuscript, as well as the anonymous referee for careful reading and detailed comments. This work was supported by the Netherlands Research Council NWO (grants #643.200.503, #639.073.803 and #614.061.608) and by the Netherlands Research School for Astronomy (NOVA). Part of the numerical computations were carried out on the Little Green Machine at Leiden University and on the LISA cluster at SURFSara in Amsterdam.

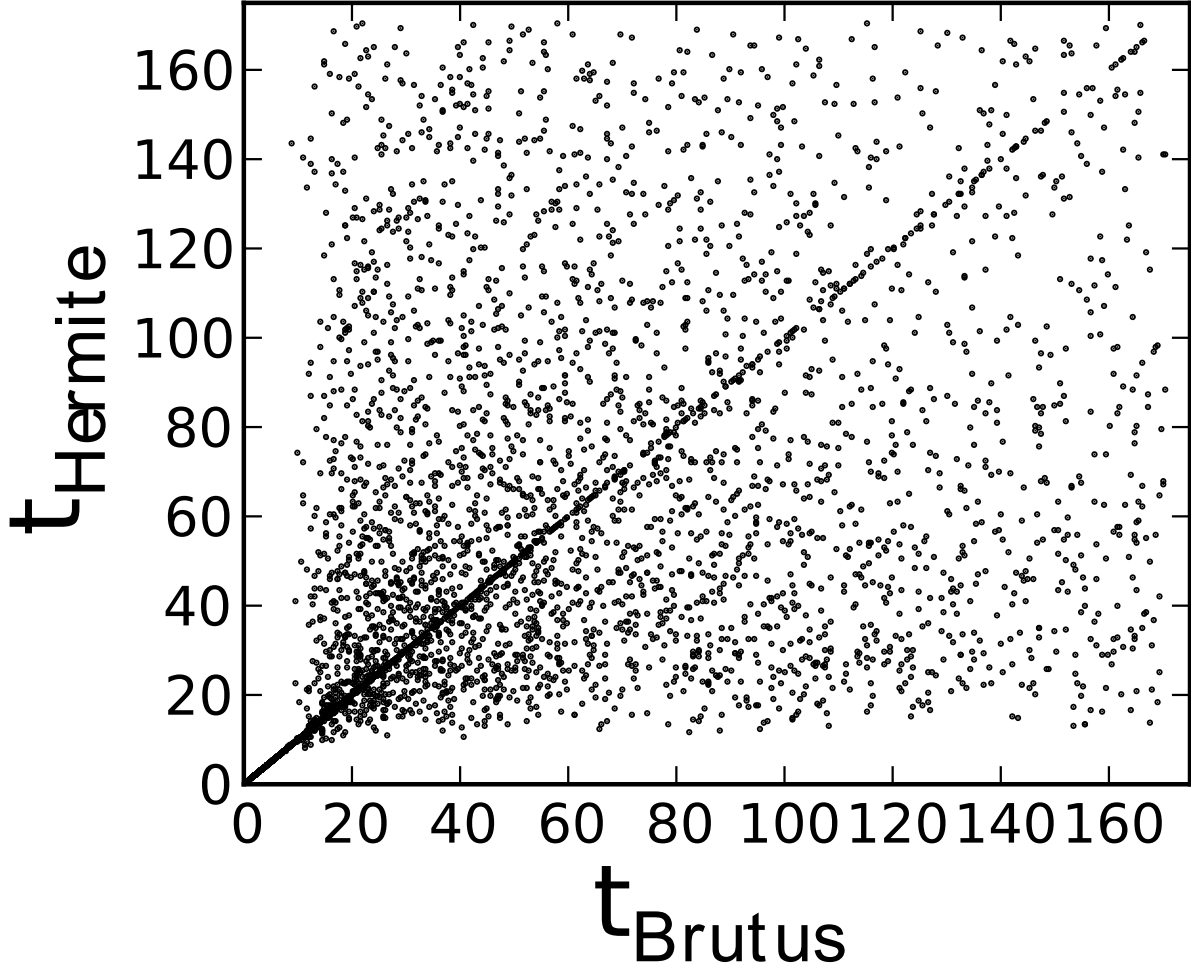


Fig. 1.— Individual comparison of the dissolution time of 3-body systems. Each point represents one unique initial realization of three equal-mass bodies taken randomly from a Plummer distribution in virial equilibrium. The time to dissolution given by Hermite (using  $\eta = 2^{-5}$ ) is on the ordinate and the converged value given by Brutus on the abscissa. About 50% of the data points lie on the diagonal which represents the cases for which Hermite and Brutus gave very similar results. The scatter around the diagonal is symmetric. For very short dissolution times ( $< 10$  dynamical times), there is insufficient time to grow errors and the results are in agreement. Once the divergence becomes important the Hermite integrator can return any value allowed in the experiment irrespective of the converged dissolution time.

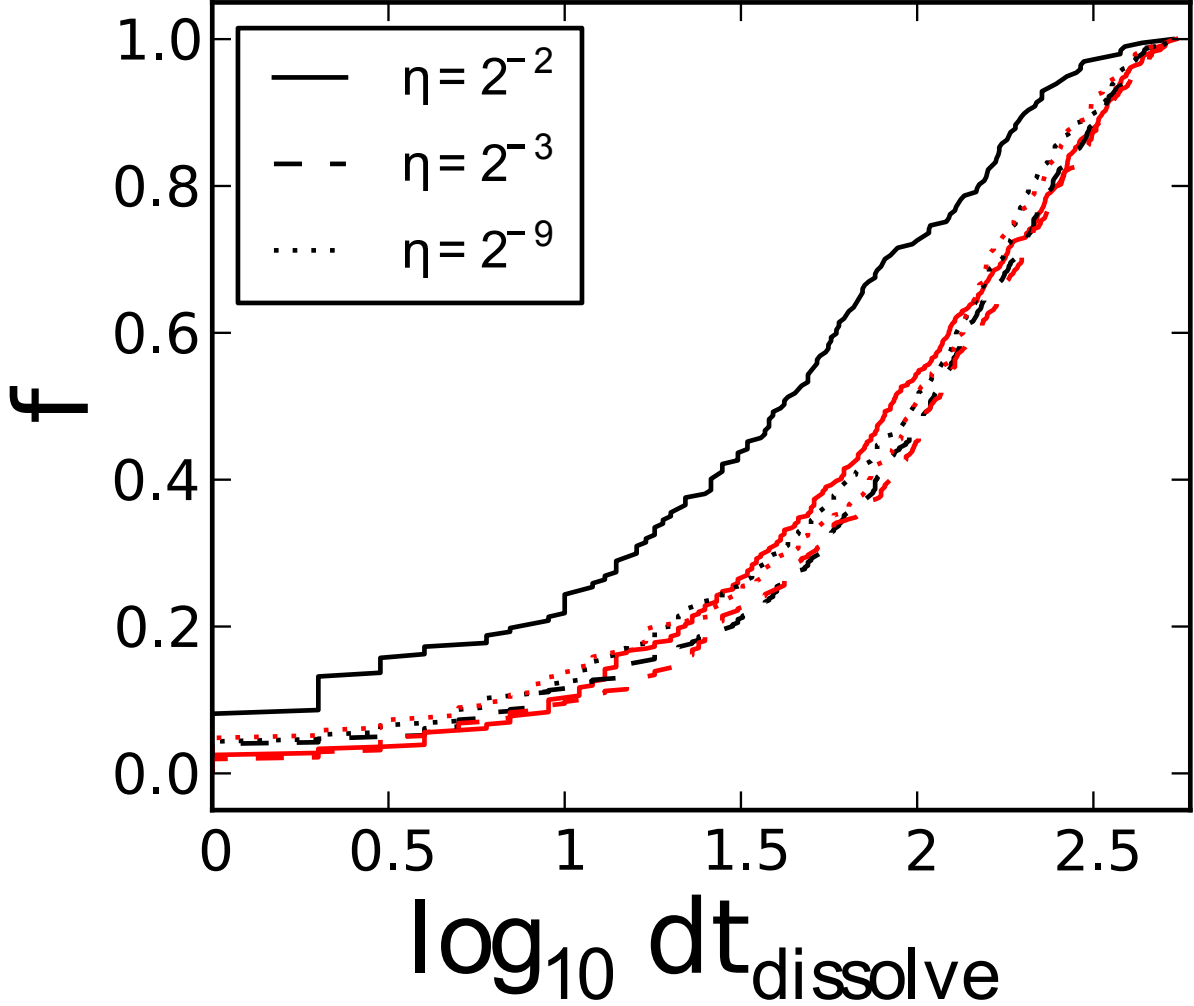


Fig. 2.— Cumulative distributions for the difference in time until dissolution for Hermite compared to converged *Brutus* solutions for three different values of  $\eta = 2^{-2}$  (solid),  $\eta = 2^{-3}$  (dashes) and  $\eta = 2^{-9}$  (dotted curve). The black curves give the distribution where the interaction calculated with Hermite lasted longer, and the red curves give the absolute value for the cases where converged solutions gave the longest lifetime for the interaction. Each (black-red) pair of curves for  $\eta \leq 2^{-3}$  is statistically indistinguishable, and the mean difference is centered around the origin.

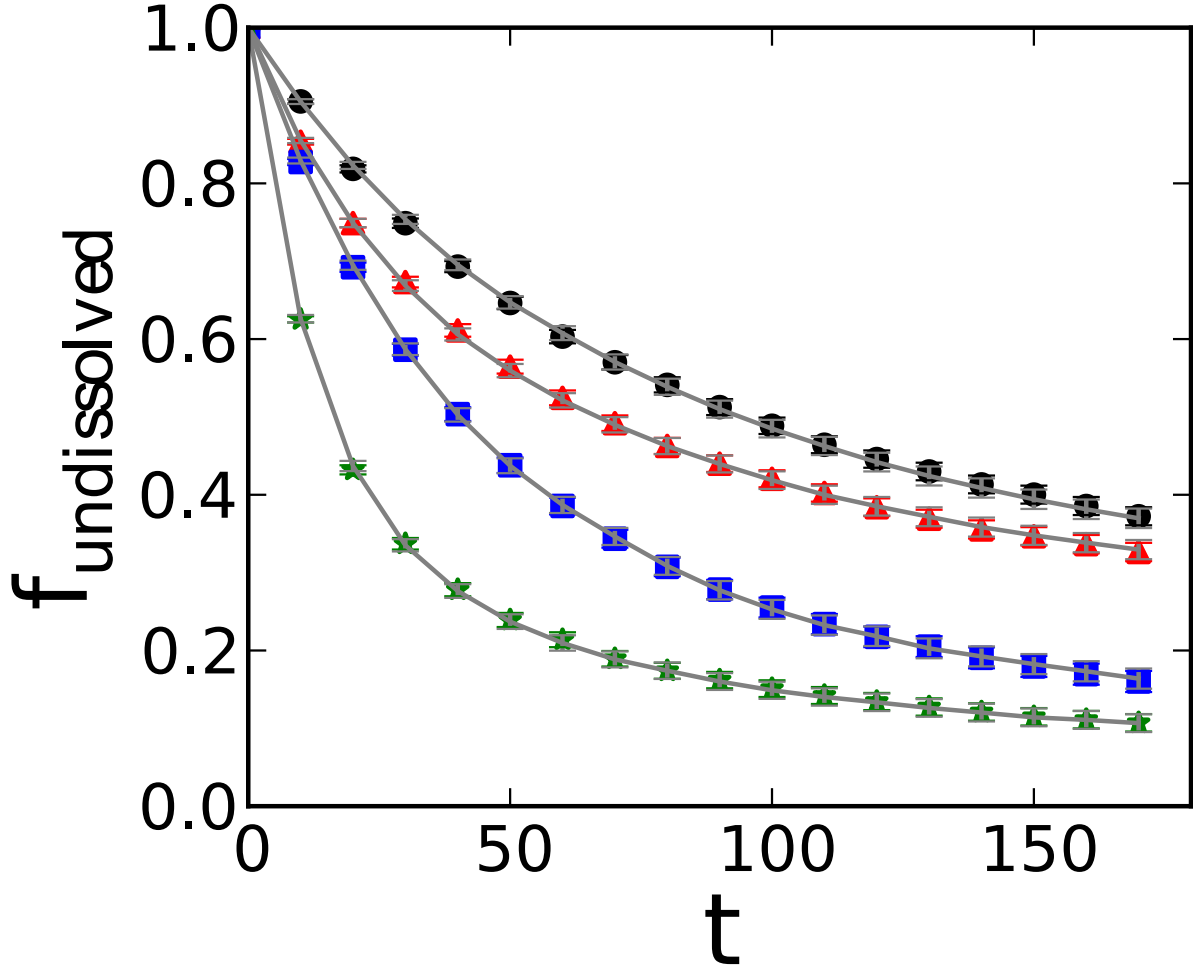


Fig. 3.— Fraction of surviving systems as a function of time for the four sets of initial conditions. The symbols give the results calculated with **Brutus**, the curves give the linear interpolation between the points calculated with the Hermite integrator using  $\eta = 2^{-5}$ . The virialized Plummer sphere with identical masses is represented by the black bullets, and with the range in masses as red triangles. The blue squares and green stars give the results for the cold Plummer distribution without and with different masses, respectively. The results of the runs with **Hermite** are statistically indistinguishable from those with **Brutus**. Up-to  $t \simeq 50$  each of the four curves can be approximated with exponential decay with a half-life time of 95, 63, 42 and 26 crossing times (from top to bottom).

## REFERENCES

- Aarseth, S. J., Lecar, M. 1975, *ARA&A* , 13, 1
- Babadzhanyants, L. K. 1981, *Pisma v Astronomicheskii Zhurnal*, 7, 752
- Barnes, J., Hut, P. 1986, *Nat* , 324, 446
- Boltzmann, L. 1868, Vol. 58, *Wiener Berichte*, 517
- Bulirsch, R., Stoer, J. 1964, *Numerische Mathematik*, 6, 413, 10.1007/BF01386092
- de Lagrange, A. M. 1772, *Essai sur le Problme des Trois Corps*
- Dejonghe, H., Hut, P. 1986, in P. Hut, S. L. W. McMillan (eds.), *The Use of Supercomputers in Stellar Dynamics*, Vol. 267 of Lecture Notes in Physics, Berlin Springer Verlag, 212
- Goodman, J., Heggie, D. C., Hut, P. 1993, *ApJ* , 415, 715
- Hayes, W. B. 2008, *MNRAS* , 386, 295
- Heggie, D., Hut, P. 2003, *The Gravitational Million-Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics*, *The Gravitational Million-Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics*, by Douglas Heggie and Piet Hut. Cambridge University Press, 2003, 372 pp.
- Heggie, D. C. 1975, *MNRAS* , 173, 729
- Hernquist, L., Hut, P., Makino, J. 1993, *ApJL* , 402, L85
- Hockney, R., Eastwood, J. 1988, *Computer Simulation Using Particles*, Adan Hilger Ltd., Bristol, UK, 540 pp.
- Hut, P., Bahcall, J. N. 1983, *ApJ* , 268, 319

- Kandrup, H. E. 1996, in R. T. Jantzen, G. Mac Keiser, R. Ruffini (eds.), Proceedings of the Seventh Marcel Grossman Meeting on recent developments in theoretical and experimental general relativity, gravitation, and relativistic field theories, p. 167
- Kepler, J. 1609, *Astronomia nova*, Vol. 1
- Lyapunov, A. M., Walker, J. A. 1994, *Journal of Applied Mechanics*, 61, 226
- Makino, J. 1994, in V. G. Gurzadyan, D. Pfenniger (eds.), *Ergodic Concepts in Stellar Dynamics*, Vol. 430 of Lecture Notes in Physics, Berlin Springer Verlag, 131
- Maxwell, J. 1867, Vol. 157, *Philosophical Transactions of the Royal Society of London*, 49
- McMillan, S., Portegies Zwart, S., van Elteren, A., Whitehead, A. 2012, in R. Capuzzo-Dolcetta, M. Limongi, A. Tornambè (eds.), *Advances in Computational Astrophysics: Methods, Tools, and Outcome*, Vol. 453 of Astronomical Society of the Pacific Conference Series, 129
- Miller, R. H. 1964, *ApJ* , 140, 250
- Monaghan, J. J. 1976, *MNRAS* , 176, 63
- Moser, J. (ed.) 1973, *Stable and random motions in dynamical systems. With special emphasis on celestial mechanics*
- Newton, I. 1687, *Philosophiae Naturalis Principia Mathematica*, Vol. 1
- Plummer, H. C. 1911, *MNRAS* , 71, 460
- Portegies Zwart, S., McMillan, S. L. W., van Elteren, E., Pelupessy, I., de Vries, N. 2013, *Computer Physics Communications*, 183, 456
- Quinlan, G. D., Tremaine, S. 1992, *MNRAS* , 259, 505

Urminsky, D. J. 2010, *MNRAS* , 407, 804

Urminsky, D. J., Heggie, D. C. 2009, *MNRAS* , 392, 1051

Valtonen, M., Mylläri, A., Orlov, V., Rubinov, A. 2004, in G. G. Byrd, K. V. Kholshevnikov, A. A. Myllri, I. I. Nikiforov, V. V. Orlov (eds.), *Order and Chaos in Stellar and Planetary Systems*, Vol. 316 of Astronomical Society of the Pacific Conference Series, 45

Verlet, L. 1967, *Phys. Rev.*, 159, 98

Vlasov, A. A. 1968, *Soviet Physics Uspekhi*, 10, 721

Wisdom, J., Holman, M. 1991, *AJ* , 102, 1528